Online ISSN 1848-3380 , Print ISSN 0005-1144
ATKAFF 52(3), 189–198(2011)

David Droeschel, Stefan May, Dirk Holz, Sven Behnke

Fusing Time-of-Flight Cameras and Inertial Measurement Units for Ego-Motion Estimation

UDK 004.93; 681.586
IFAC 2.8; 4.2.2

Original scientific paper

This paper presents an approach for ego-motion estimation based on a Time-of-Flight (ToF) camera. It is applied in real-time on a mobile robot platform during mission.

The proposed method utilizes the coherence of depth and reflectance data of ToF cameras by detecting image features on reflectance data and estimating the motion on related depth data. This motion estimate is fused with data from an inertial measurement unit in order to gain higher accuracy and robustness, especially in situations when image registration fails.

Results are benchmarked against reference poses from an accurate laser ranger finder-based localization.

Key words: Ego-Motion estimation, ToF camera, Sensor fusion

Fuzija TOF kamera i inercijalnih mjernih jedinica za procjenu vlastitog gibanja. Članak opisuje pristup procjeni vlastitog gibanja utemeljenom na time-of-flight (ToF) kamerama. Metoda je primijenjena u stvarnom vremenu na mobilnom robotu tijekom misije. Predložena je metoda utemeljena na slaganju dubinskih i reflektiranih podataka iz ToF kamera, detektira značajke u slici reflektiranih podataka te procjenjuje gibanje pripadajućih dubinskih podataka. Vrši se fuzija procjene gibanja s podacima iz inercijalne mjerne jedinice radi ostvarivanja bolje točnosti i robusnosti, pogotovo u slučaju nemogućnosti registracije slike. Rezultati su uspoređeni s referentnim mjerenjima algoritma lokalizacije temeljenog na točnom laserskom senzoru udaljenosti.

Ključne riječi: procjena vlastitog gibanja, ToF kamera, fuzija senzora

1 INTRODUCTION

Time-of-flight (ToF) cameras are compact, solid-state sensors that provide depth and reflectance measurements at high frame rates. An array of infrared LEDs is employed to illuminate the environment with modulated light. The reflected light is measured on a photo-sensitive chip in CCD/CMOS technology, i.e., depth and intensity are determined in parallel for each element in the sensor array.

The dense depth measurements facilitate 6-DoF registration of frame sequences. However, accuracy and precision of the distance measurements are limited by a number of error sources. Therefore, range image registration is error prone and results in highly distorted maps even in environments of small scale. Detailed explanations on the working principle and related error sources can be found in the works of Lange [1] and Schneider [2].

ToF cameras have some advantages when comparing them to other state-of-the-art 3D sensors. So far, 3D laser scanners and stereo camera systems are mostly used due to their high measurement range and accuracy.

Stereo vision induces a high computational load since correspondences between two images from a different perspective have to be determined. In addition, distances to structureless surfaces cannot be measured, if the perspective projection of the object is larger than the camera's field of view. Stereo vision also suffers from shadowing effects and changes in illumination. These issues can be avoided using 3D laser range finders that actively illuminate the scene. 3D laser range finders have other disadvantages though, like high power consumption, mechanical moving parts and low frame rate.

The work presented in this paper utilizes a ToF camera for ego-motion estimation. It incorporates the coherent reflectance and depth images measured by the camera. Point correspondences are determined directly on the low-resolution reflectance data. The depth information is incorporated for estimating the transformation between frames. In order to increase robustness, the motion estimate is fused with the measurements of an inertial measurement unit (IMU). Both sensors are complementary and, thus, can compensate the errors of each other. Range image registration with a narrow field-of-view has difficulties in situations with high rotational velocities, whereas the transla-

Fig. 1. The resulting 3D map based on the estimated ego-motion.

tional uncertainties of the IMU can be compensated while performing only translational motion in forward direction.

The remainder of the paper is structured as follows: Section 2 summarizes the related work. Section 3 describes necessary pre-processing steps. Section 4 and 5 present our main contributions: an approach to estimate the ego-motion from ToF camera data and a model for fusing it with inertial measurements, respectively. Section 6 reports on experiments that have been carried out to evaluate the proposed method.

2 RELATED WORK

The first application of ToF cameras in mobile robotics was published in 2004. Weingarten et al. [3] used the Swissranger SR-2 device for basic obstacle avoidance and local path planning. Their experiments showed that path planning and obstacle avoidance based on the ToF camera measurements could prevent the robot from colliding with an obstacle that was not detected by the 2D laser range finder. Sheh et al. [4] used a ToF camera for 3D mapping of a RoboCup Rescue environment. Because of the narrow apex angle, they rotated the camera on a pan-tilt unit to gain a larger field of view. The robot stopped at every location and took 10 range images at different pan-tilt positions. The acquisition of one scan took 20 seconds and registration of the range images was assisted by a human operator. Ohno et al. [5] also used a SR-2 device to estimate the robot's ego-motion by means of 3D range image registration based on the *Iterative Closest Point* (ICP) algorithm. The resulting trajectory was compared to a reference trajectory. The experiments involved almost straight trajectories with up to 6.5 m distance. The authors mentioned that in larger scenes with less structure the rotational error would be higher and that the use of a gyroscope could compensate this error. They mainly used algorithms that have been successfully applied to laser range finder data. Applying these methods to the ToF camera is not straightforward mainly for two reasons: Bauer

- Compared to laser ranger finders, the measurement accuracy of recent ToF cameras is lower.
- Due to the larger field of view of laser range-finders, the registration of range images is easier.

Because of the lower measurement accuracy of ToF cameras, many groups addressed error modeling and calibration. Lindner et al. [6] as well as Kahlmann et al. [7] estimated intrinsic parameters of a ToF camera using the reflectance image of a checkerboard and a planar test field with Near-Infra-Red (NIR) LEDs, respectively. A per-pixel precision of at least 10mm was achieved.

Regarding the registration of range images, the ICP algorithm [8] is the most popular approach. It iteratively estimates the transformation between two point clouds, the *model point set* and the *scene point set*. In every iteration, the point correspondences between model and scene are determined by a nearest neighbor search and the transformation between the point correspondences is estimated by a least squares minimization. The mean squared error of the estimated transformation applied to the scene is determined in every iteration. The algorithm continues until the error converges or a maximum number of iterations is reached. There are many variations of the ICP algorithm. The application of the ICP to ToF camera measurements has been studied by May et al. [9]. A practical problem in the application of the ICP algorithm is the convergence to local minima. This is particularly the case in scenes with low geometric structure. These situations occur more often when having a smaller field of view, e.g., for ToF cameras. Sheh et al. [4] handled this problem by using a pan-tilt unit which results in a low data acquisition rate. In scenes where the geometric structure is low but the texture of the objects is high, image features from the reflectance image of the camera could contribute to a better motion estimate.

Swadzba et al. [10] performed only the coarse registration based on depth of features in the reflectance image, whereas the fine registration has been calculated by the use of the whole depth image. To compensate for the problems of ToF cameras, researchers proposed the fusion with other sensors. Prusak et al. [11] presented a joint approach for robot navigation with collision avoidance, pose estimation and map building employing a ToF camera combined with a high-resolution spherical camera. A rough registration was performed on the circumferential view of the spherical camera. The registration was then refined using the range

image. Huhle et al. [12] also used a joint approach of feature tracking and range image registration. They fused images from a high-resolution color camera with ToF camera measurements to obtain colored 3D point clouds. Besides that, they incorporate the information of an IMU to initialize registration. Feature determination performed much better on the high-resolution color images than on the low-resolution monochromatic reflectance images of the ToF device. A comparison of tracking based on measurements acquired from a ToF camera with images acquired using a high-resolution color camera has been provided by Sabeti et al. [13]. They concluded that high-resolution color sensors are more suitable for outdoor applications and purposes that require the detection of fine details while ToF sensors are more appropriate for object tracking that requires information about the distance between object and camera.

3 PRE-PROCESSING

The accuracy of measurements with ToF cameras in unknown scenes varies considerably, due to error effects inherent to their functional principle. We apply a set of preprocessing steps to discard and correct noisy and erroneous measurements.

3.1 Jump Edge Filtering

In a first step, points at so-called jump edges are filtered out. Jump edges occur when a foreground object occludes a background object. The true distance changes suddenly at the transition from foreground to background, but ToF cameras measure a smooth transition. Details on the used jump edge filter can be found in a previous publication [9]. Since the jump edge filter is sensitive to noise, we apply a median filter to the distance image beforehand.

3.2 Amplitude Filtering

Noise errors can be rated with respect to the amplitude measurements, since the accuracy of distance measurements increases with the amount of light returning to the sensor. Thus, measurements with a low amplitude (below a certain threshold) should be neglected. A pure amplitude based filtering would be disadvantageous in terms of narrowing the field of view. Due to the inhomogeneous image illumination, measurements in the peripheral part of the field of view would be discarded more often. Thus, amplitude filtering is a trade-off between discarding noisy measurements and discarding measurements in the peripheral part of the field of view. Lowering the threshold sorts out more erroneous measurements but also further narrows the field of view.

3.3 Phase Unwrapping

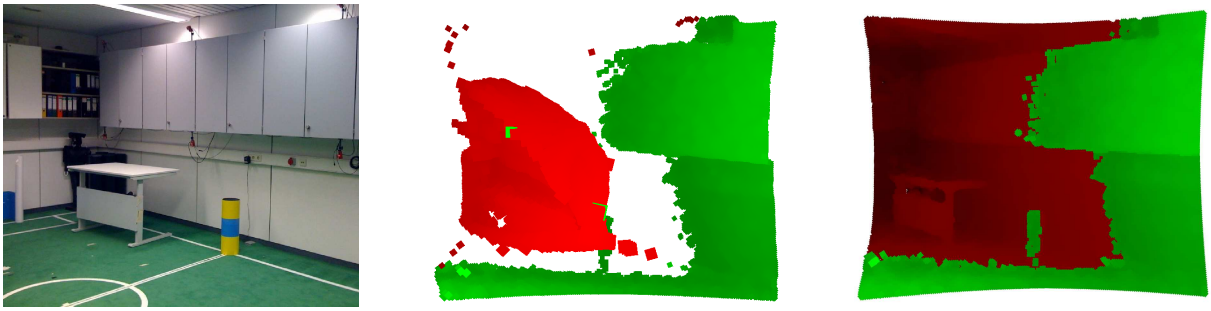
ToF cameras gain depth information by measuring the phase shift between emitted and reflected light, which is proportional to the object's distance modulo the wavelength of the modulation frequency. This results in a distance ambiguity: measurements beyond the sensor's non-ambiguity range are wrapped into the non-ambiguity range and result in artifacts and spurious distance measurements. A common way to handle these distance ambiguities is to neglect measurements based on the ratio of the measured distance and amplitude, since the amplitude of the reflected signal decreases with the square of the distance to an object. The drawback of this method is that information is lost due to neglecting measurements. Another limitation of this approach is that wrapped measurements are only detected by the ratio of distance and amplitude, not taking into account the gradient surface of neighboring measurements, which results in wrong classifications for distant objects with high infrared reflectivity.

We correct wrapped distance measurements by identifying a number of *phase jumps* in the distance image, i.e., the relative wrappings between every pair of neighboring measurements. We use a probabilistic approach [14] that detects discontinuities in the depth image to infer these phase jumps. The application of phase unwrapping is shown in Fig. 2.

4 EGO-MOTION ESTIMATION

To estimate the camera's motion between two consecutive frames, image features in the reflectance image of the ToF camera are used to determine point correspondences between the frames. To detect image features, the Scale Invariant Feature Transform (SIFT) [15, 16] is used. SIFT features are invariant in rotation and scale and are robust against noise and illumination changes. The SIFT algorithm has been shown to outperform other feature extraction methods [17]. Various refinements of the basic SIFT algorithm have been proposed, e.g. PCA-SIFT [18], GLOH [19], and SURF [20]. Bauer et al. [21] compare recent implementations of SIFT and SURF. They show that SIFT yields the best results regarding the *ratio* of incorrect and correct matches and the total number of correct matches.

In order to estimate the camera motion between two frames, the features of one frame are matched against the features of the other frame. As described in [16], the best match is the nearest neighbor in the 128-dimensional keypoint descriptor space. To determine the nearest neighbor, the Euclidean distance is used. To measure the quality of a match, a distance ratio between the nearest neighbor and the second-nearest neighbor is considered. If they are too similar, the match is rejected. Hence, only features that



(a) (b) (c)

Fig. 2. Phase unwrapping of an indoor scene. (a) image of the scene. (b) + (c) 3D point clouds that have been generated based on the camera's depth image. Color of the points indicates the result of the algorithm; wrapped measurements are shown in red. Brightness encodes distance to the camera center. (b) point cloud without unwrapping. Measured distances beyond the sensor's non-ambiguity range are wrapped into it, which results in artifacts between 0 and 3 meters. (c) unwrapped depth image.

(a) (b) (c)

Fig. 3. SIFT feature extraction and matching applied on a ToF reflectance image. The scene shows a robot in the pavilion at the Fraunhofer IAIS. Images (a) and (b) show the detected SIFT features on two consecutive camera frames. The number of detected features are 475 (a) and 458 (b). (c) matching result: 245 features from image (a) are matched to features from image (b). White lines indicate feature displacement.

are unambiguous in the descriptor space are considered as matches. Experiments have shown that a distance ratio of 0.6 results in the best rejection rates in our case.

Figure 3 shows the reflectance images of two consecutive frames. The red and green dots show detected features in both images. Figure 3 (c) shows the matching results of the two images. The green dots are the features from image (a) and the red dots are the matched features from frame (b). 245 features are successfully matched.

Each match constitutes a point correspondence between two frames. By knowing the depth of every pixel, a point correspondence in 3D is known. The set of points from the current frame is called the *data* set, and the set of corresponding points in the previous frame is called the *model* set. The scene is translated and rotated by the sensor's ego motion. Thus, the sensor ego-motion can be deduced by finding the best transformation that maps the data set to the model set. We employ the method for estimating the rigid transformation proposed by Arun et al. [22]. It uses a closed form solution for estimating the 3×3 rotation matrix $\mathbf{R}$ and the translation vector $\vec{t}$, which is based on singular value decomposition (SVD).

The distances between corresponding points, after applying the estimated transformation, forms the *Root Mean Square* (RMS) error. The RMS error is often used in range registration to evaluate the scene-to-model consistency. It can be seen as a measure for the quality of the match: if the RMS error is high, the scene-to-model registration cannot be consistent. On the other hand, a low RMS error does not imply a consistent scene-to-model registration, since it also depends on the number and distribution of the point correspondences.

The translation vector $\vec{t}$ is composed of $(\Delta x, \Delta y, \Delta z)^T$, which is the translational change of the camera between two frames. The rotation matrix $\mathbf{R}$ is the change of the camera orientation between two frames. From the rotation

matrix the three Euler angles can be calculated.

Since our robot is moving on planar ground, the pose estimate can be simplified. Hence, the translation $(\Delta x, \Delta y)^T$ and the rotation around the vertical (yaw) axis $\Delta\theta$ can be considered as the transformation that describes the camera's motion between two frames.

From the $(\Delta x_k, \Delta y_k, \Delta\theta_k)^T$ at frame k, the trajectory of the camera can be built incrementally. The pose $(x_k, y_k, \theta_k)^T$ at frame k can be calculated by

$$(x_k, y_k)^T = (x_{k-1}, y_{k-1})^T + \mathbf{R}(\Delta\theta_k)(\Delta x_k, \Delta y_k)^T \quad (1)$$

and

$$\theta_k = \theta_{k-1} + \Delta\theta_k, \quad (2)$$

where $k-1$ is the previous frame and $\mathbf{R}(\Delta\theta_k)$ is the 2D rotation matrix of θ_k.

5 FUSION OF MOTION ESTIMATES

The motion estimation described in the previous section provides a translational $(\Delta x, \Delta y)^T$ and rotational change $\Delta\theta$ of the camera between two camera frames. By knowing the time between two frames Δt, the translational and rotational velocity is known. This is considered as observation

$$z_k = \frac{1}{\Delta t}(\Delta x, \Delta y, \Delta\theta)^T \quad (3)$$

at time step $\mathbf{k}$.

A second source of motion measurements is a XSens MTi inertial measurement unit, calibrated with the vendor's calibration toolbox. It provides measurements for the rotational velocity v_θ and translational acceleration on the x and y axis, $(\vec{a}_x, \vec{a}_y)$. The *mean* rotational acceleration $\vec{a}_{\theta,k}$ at time step k can be calculated by the difference between the velocities at time steps k and $k-1$:

$$\vec{a}_{\theta,k} = v_{\theta,k} - v_{\theta,k-1}. \quad (4)$$

A Kalman filter predicts the system velocity estimate $(v_x, v_y, v_\theta)^T$. The motion estimate from the camera is considered as observation, whereas the IMU data is considered as control input to the system. The RMS error and its individual components of the estimated transformation reflect the certainty of the observation and are therefore used as an approximation of the observation covariance (similar to [23]).

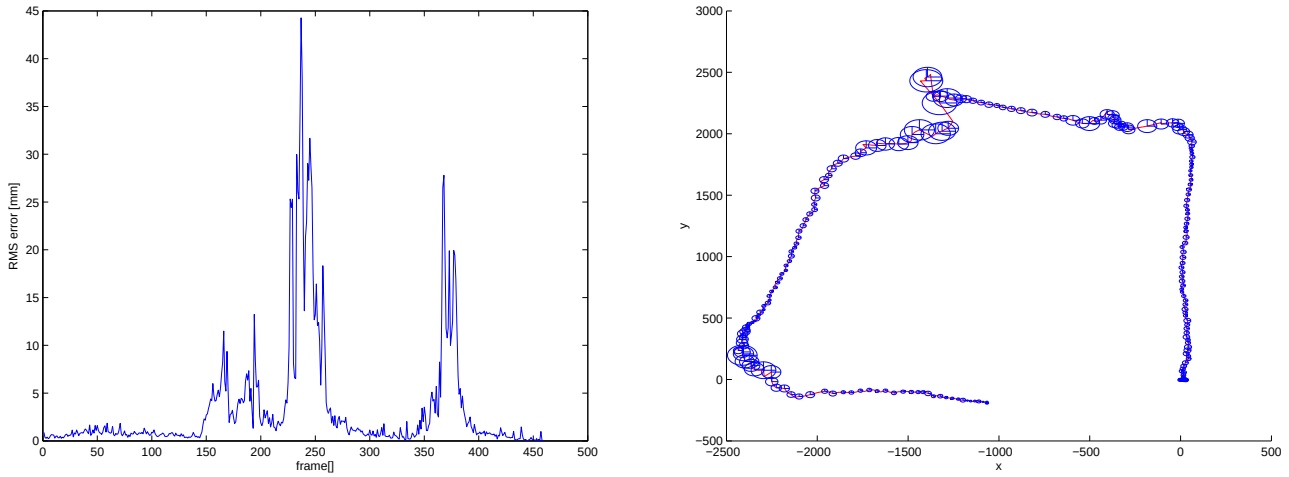
Fig. 4. The scene of the first experiment carried out in the Robotik Pavillion at Fraunhofer IAIS, Sankt Augustin. The scene consists of a wooden staircase with a robot on it, some posters and a calibration pattern. The robot moved on a square with $120\,\mathrm{cm}$ *side length.*

6 EVALUATION

The following experiments demonstrate the accuracy and robustness of the proposed procedure. The employed sensor is a ToF camera, the SR3000 by Mesa Imaging. A Sick LMS200 laser range finder was used to incrementally construct an accurate and consistent 2D map. The ICP algorithm is applied to the laser-range scans to generate a reference trajectory.

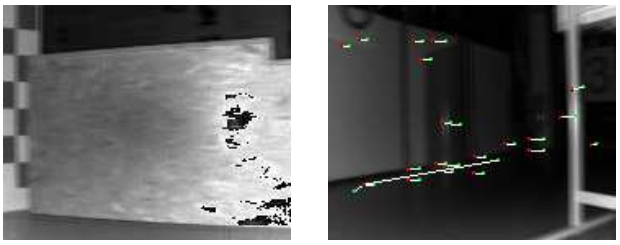
Figure 4 shows the scene of the first experiment. The experiment was carried out in the Robotic Pavilion at Fraunhofer IAIS, Sankt Augustin. The image shows a wooden staircase with a robot, some posters and a calibration pattern. The robot moved along a square with $120\,\mathrm{cm}$ side length. Figure 6(a) compares the trajectory estimated from the ToF camera alone and the trajectory estimated from ToF and IMU to the reference trajectory. The fused trajectory is less distorted than the ToF-only trajectory, especially in situations where the RMS error is high, e.g., in curves.

Figure 5(a) depicts the RMS error of the estimated transformation applied to the matched point pairs. The first 150 frames show a relative low RMS error, compared to the peak at frame 245. To visualize the correlation of the RMS error and the distorted trajectory, Fig. 5(b) shows the estimated trajectory as well as the RMS error distribution. The figure shows high RMS errors at those poses that deviate from the reference trajectory. Fig. 7 shows the camera's amplitude image at two frames where the RMS error is large. One can see that at frame 160 the image is distorted due to the motion blurring effect and at frame 245



(a) (b)

Fig. 5. (a) RMS error of the estimated transformation applied to the matched point pairs for each frame. The first 150 frames show a relative low RMS error compared to the peak at frame 245. (b) The trajectory is plotted by a red line and the RMS error is visualized by blue ellipses, where the magnitude of the RMS error correlates to the size of the ellipsis.



(a) (b)

Fig. 7. ToF camera amplitude image at two frames where the RMS error is large. At frame 160 (a) the image is distorted due to the motion blurring effect and at frame 245 (b) the number and the distribution of features is low.

the number and the distribution of features is low.

Figure 8 shows the translational error, the rotational error for every frame and the cumulated rotational error for the unfiltered and the filtered motion estimate. The Kalman-filtered motion estimate improves up to $1006\,\mathrm{mm}$ on the translational error and up to $25.4°$ on the rotational error.

Figure 6(b) shows the estimated trajectories of a second experiment. In a larger scene with up to $8\,\mathrm{m}$ diameter. Figure 9 depicts the translational and rotational error of the applied methods, comparing the ego-motion estimate based solely on the camera data to the fused ego-motion estimate. The rotational error of the fused ego-motion estimate improves up to $28.6°$. The improvement of the translational error is up to $669\,\mathrm{mm}$.

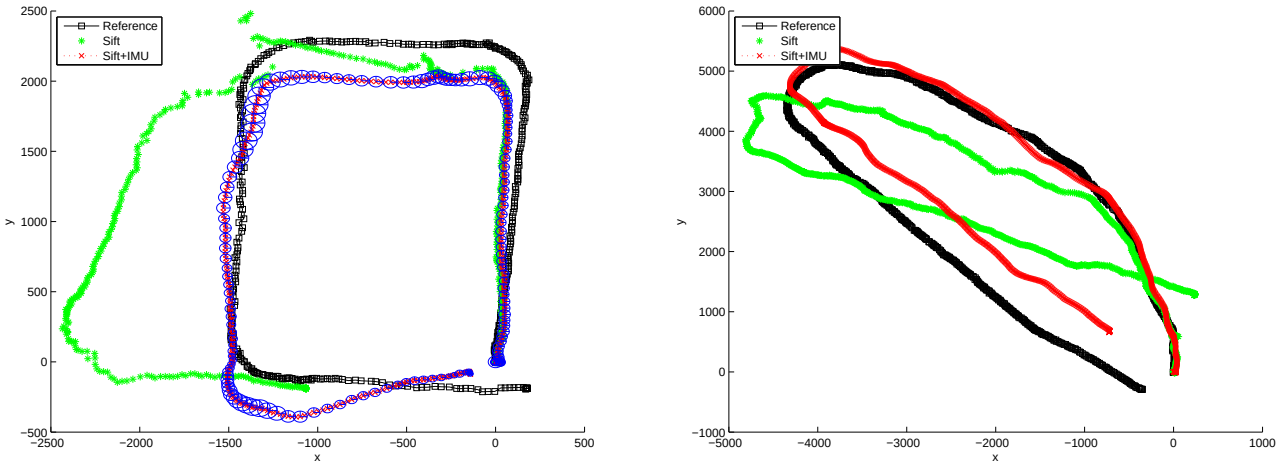
Figure 10 shows the resulting 3D maps based on the estimated ego-motion. Figure 6 depicts the unfiltered motion estimate. The resulting map is squeezed at the end of the trajectory due to the error in the ego-motion estimate. In contrast, Figure 6 shows the improved map based on the fused ego-motion estimate.

With the camera's relative low resolution of 176×144, we achieve an average runtime of $102\,\mathrm{ms}$ per frame for extracting the keypoints in the current frame and matching them against the keypoints from the previous frame on a $1.66\,\mathrm{GHz}$ single core of a Core2Duo laptop computer. However, the runtime depends on the number of feature points, which corresponds to the amount of texture in the scene. The extraction and the matching could be parallelized.

7 CONCLUSIONS

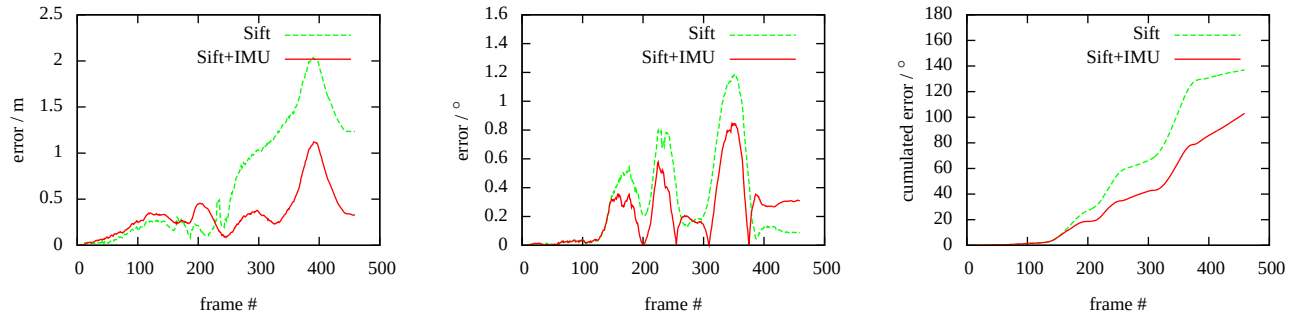
This paper presented a way to estimate a robot's ego-motion while moving. An application of this motion estimate is to map an unknown environment based on data from a ToF camera. ToF cameras provide depth and reflectance data of the scene at a high frame rate. They suffer from a set of error sources which make them difficult to handle. The proposed method utilizes the coherence of depth and reflectance data of ToF cameras by detecting image features on the reflectance data and estimating the motion on the depth data.

The visual motion estimate is fused with the IMU measurements to gain higher accuracy and robustness. The evaluation shows that fusing the pose estimate with the data from the IMU improves the estimated trajectory.



(a) (b)

Fig. 6. (a) Estimated trajectories. The black trajectory shows the reference data based on the 2D laser range finder. The green trajectory shows the SIFT-based motion estimate. On the upper left corner, the trajectory is distorted. The application of the sensor fusion is depicted in the red trajectory. The blue ellipses on the red trajectory depict the a posteriori system covariance of the Kalman filter. (b) Estimated trajectories of a second experiment.



(a) (b) (c)

Fig. 8. (a) Translational error of the unfiltered (dashed green) and Kalman-filtered (red) motion estimate compared to the reference from the 2D laser range finder. (b) Rotational error. (c) Cumulated rotational error.

The fusion of IMU and range image registration based transformations also increased robustness. For large rotational velocities, the error in registration is high due to a smaller overlap. In this situation, an IMU is valuable. On the other hand, performing a pure translation is handled properly by range image registration. The ToF-based motion estimate also prevents drift. A mechanism to deal with those situations is topic for future work.

In this work, the robot was moving on a planar ground, and a 2D laser range finder was used as reference. Future work will investigate in the extension to full 6D poses.

Another important point is the determination of the observation covariance. Here, the RMS error was used as a first attempt. In the future work, a camera-specific error model has to be considered.

ACKNOWLEDGMENT

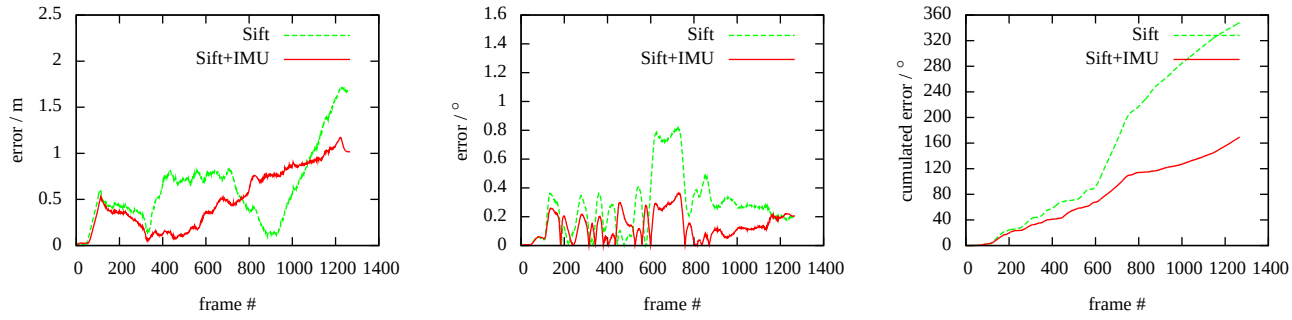
The continued support by the b-it foundation is gracefully acknowledged.

REFERENCES

[1] R. Lange, *3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology*. PhD thesis, University Siegen, 2000.

[2] B. Schneider, *Der Photomischdetektor zur schnellen 3D-Vermessung für Sicherheitssysteme und zur Informationsübertragung im Automobil*. PhD thesis, Universität-Gesamthochschule Siegen, 2003.

[3] J. Weingarten, G. Gruener, and R. Siegwart, "A state-of-the-art 3D sensor for robot navigation," in *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, vol. 3, pp. 2155 – 2160, 2004.



(a) (b) (c)

Fig. 9. (a) Translational error of the unfiltered (dashed green) and filtered (red) motion estimate compared to the reference data from the 2D laser range finder. (b) Rotational error. (c) Cumulated rotational error.

[4] R. Sheh, M. W. Kadous, and C. Sammut, "On building 3D maps using a range camera: Applications to rescue robotics," techreport, ARC Centre of Excellence for Autonomous Systems - The University of New South Wales, Sydney Australia, April 2006.

[5] K. Ohno, T. Nomura, and S. Tadokoro, "Real-time robot trajectory estimation and 3D map construction using 3D camera," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pp. 5279 –5285, 2006.

[6] M. Lindner and A. Kolb, "Lateral and depth calibration of pmd-distance sensors," in *Advances in Visual Computing*, vol. 2, pp. 524–533, Springer, 2006.

[7] T. Kahlmann, F. Remondino, and H. Ingensand, "Calibration for increased accuracy of the range imaging camera swissranger," in *Proceedings of the ISPRS Commission V Symposium 'Image Engineering and Vision Metrology'*, vol. XXXVI, (Dresden, Germany), pp. 136–141, 2006.

[8] P. Besl and N. McKay, "A method for Registration of 3–D Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[9] S. May, D. Droeschel, D. Holz, S. Fuchs, E. Malis, A. Nüchter, and J. Hertzberg, "Three-dimensional mapping with time-of-flight cameras," *Journal of Field Robotics, Special Issue on Three-Dimensional Mapping, Part 2*, vol. 26, no. 11-12, pp. 934–965, 2009.

[10] A. Swadzba, B. Liu, J. Penne, O. Jesorsky, and R. Kompe, "A comprehensive system for 3D modeling from range images acquired from a 3D ToF sensor," in *The 5th International Conference on Computer Vision Systems*, 2007.

[11] A. Prusak, O. Melnychuk, H. Roth, I. Schiller, and R. Koch, "Pose estimation and map building with a PMD-camera for robot navigation," in *Proceedings of the Dynamic 3D Imaging Workshop in Conjunction with DAGM (Dyn3D)*, (Heidelberg, Germany), 2007.

[12] B. Huhle, P. Jenke, and W. Straßer, "On-the-fly scene acquisition with a handy multisensor-system," in *Proceedings of the Dynamic 3D Imaging Workshop in Conjunction with DAGM (Dyn3D)*, (Heidelberg, Germany), 2007.

[13] L. Sabeti, E. Parvizi, and Q. J. Wu, "Visual Tracking Using Color Cameras and Time-of-Flight Range Imaging Sensors," *Journal of Multimedia*, vol. 3, pp. 28–36, 2008.

[14] D. Droeschel, D. Holz, and S. Behnke, "Probabilistic phase unwrapping for time-of-flight cameras," in *Proceedings of the joint conference of the 41st International Symposium on Robotics (ISR 2010) and the 6th German Conference on Robotics (ROBOTIK 2010), Munich, Germany*, 2010.

[15] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2, pp. 1150 –1157 vol.2, 1999.

[16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[17] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," in *Int. Conference on Computer Vision & Pattern Recognition*, vol. 2, pp. 257–263, 2003.

[18] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 2, pp. 506–513, 2004.

[19] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Tr. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[20] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded-up robust features," in *9th European Conference on Computer Vision*, (Graz, Austria), 2006.

[21] J. Bauer, N. Sünderhauf, and P. Protzel, "Comparing several implementations of recently published feature detectors.," in *Proceedings of the International Conference on Intelligent and Autonomous Systems, IAV*, 2007.

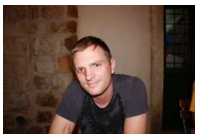
[22] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Tr. on Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 698–700, 1987.

[23] A. J. Stoddart, S. Lemke, A. Hilton, and T. Renn, "Estimating Pose Uncertainty for Surface Registration," in *Proceedings of British Machine Vision Conference (BMVC)*, 1996.

(a)

(b)

Fig. 10. (a) Top view of the resulting 3D map based on the estimated ego-motion. The map is squeezed at the end of the trajectory due to the error in the ego-motion estimate. (b) Improved map based on the fused ego-motion estimate.



David Droeschel received a M.Sc. degree in Autonomous Systems from the University of Applied Sciences Bonn-Rhein-Sieg in 2009. Since May 2009, he is a researcher in the Autonomous Intelligent Systems Group of the University of Bonn and a member of the RoboCup@Home team NimbRo. His research interests include 3D perception, human-robot interaction, and people perception.

Stefan May received a diploma (electrical engineering) in 2000 and a Master's degree (software engineering) in 2004 from the Georg Simon Ohm University of Applied Sciences in Nuremberg. In 2009 he received his PhD in computer vision from the University in Osnabrueck. Between 2002 and 2009 he was working in the context of automotive electronics and 3D vision for mobile robotics at Audi and Fraunhofer IAIS in Germany and INRIA in France. Since 2010, he is professor for Automation and Mechatronics at the Georg Simon Ohm University of Applied Sciences in Nuremberg. His research interests include sensors and data processing for 3D perception of mobile agents.

Dirk Holz received a Diploma in Computer Science from the University of Applied Sciences Cologne in 2006 and a M.Sc. degree in Autonomous Systems from the University of Applied Sciences Bonn-Rhein-Sieg in 2009. Since May 2009, he has been working as a researcher in the Autonomous Intelligent Systems Group of the University of Bonn. He has been participating in RoboCup@Home since 2007 and is a member of the league's Technical Commitee since 2009. His research interests include 3D perception and world modeling.

Sven Behnke received a Diploma in Computer Science in 1997 from Martin-Luther-Universität Halle-Wittenberg and PhD from Freie Universität Berlin in 2002. He spent the year 2003 as postdoctoral researcher at the International Computer Science Institute, Berkeley, CA. From 2004 to 2008, Professor Behnke headed the Humanoid Robots Group at Albert-Ludwigs-Universität Freiburg. Since April 2008, he is professor for Autonomous Intelligent Systems at the University of Bonn. His research interests include humanoid robots, computer vision, and machine learning.

AUTHORS' ADDRESSES

David Droeschel
Dirk Holz
Prof. Dr. Sven Behnke
Autonomous Intelligent Systems Group
Institute of Computer Science
Rheinische Friedrich-Wilhelms-Universität Bonn
Friedrich-Ebert-Allee 144
53113 Bonn
email: behnke@cs.uni-bonn.de, {droeschel, holz}@ais.uni-bonn.de

Prof. Dr. Stefan May
Department of Electrical Engineering, Precision Engineering, Information Technology
Georg Simon Ohm University of Applied Sciences Nuremberg
Keßlerplatz 12
90489 Nürnberg
email: Stefan.May@ohm-hochschule.de

Received: 2010-03-02
Accepted: 2011-07-14